



Abstract ampliado

RESUMEN AMPLIADO

Título: Limitaciones del uso de investigaciones pre-siglo XXI como referencia en estudios de renta y desigualdad

Autores y e-mail de todos ellos:

¹M^a Isabel López Rodríguez: Maria.I.Lopez@uv.es

²Daniel Gonzalo Palací-López: Dapalpe@gmail.com

³Jesús Palací: Jpalaci@comillas.edu

Departamentos:

¹Economía Aplicada

²Departamento de Estadística e I.O. Aplicadas y Calidad

³Departamento de Electrónica, Automática y Comunicaciones

Universidades:

¹Universitat de València

²Universitat Politècnica de València

³Universidad Pontificia Comillas ICAI.

Área Temática: *Movimientos de la población, desigualdad, pobreza y crecimiento inclusivo*

Resumen:

Introducción: La importancia económica y social del estudio de la distribución de la renta y de la desigualdad es innegable, constituyendo uno de los objetivos prioritarios de los poderes gubernamentales. En el caso de España y desde finales del siglo pasado hasta la actualidad se ha utilizado mayoritariamente, para los estudios de renta, la información empírica aportada por las Encuestas de Presupuestos Familiares (E.P.F.).

El análisis de la evolución de la desigualdad, tan en boga en la coyuntura económica actual, requiere una comparativa en la que se toma como referencia las conclusiones obtenidas en las investigaciones que han hecho uso de dicha información.

Cabe, por tanto, preguntarse si el referente es adecuado, es decir, si hubo y en qué sentido una repercusión de la naturaleza de los datos en las conclusiones obtenidas.

Dicho objetivo surge ante la heterogeneidad del formato de dichos datos, agregados en decilas en las primeras encuestas y desagregados en las siguientes. La disponibilidad de los datos agregados y desagregados correspondientes a las E.P.F. de 1980-81 y 1990-91, permitirá realizar el análisis propuesto.

Para ello resulta necesario, además, modelizar haciendo uso de dichos datos la renta.

En ese sentido cabe reseñar que desde los estudios pioneros de Pareto hasta nuestros días, son múltiples los trabajos cuya finalidad ha sido identificar el modelo que permita describir la renta de los hogares.

Sin ánimo de realizar un estudio exhaustivo de los mismos, cabe mencionar, además de la precursora distribución de Pareto, la distribución de Champernowne, debida al autor con igual nombre, las distribuciones de Dagum (tipo I, tipo II y tipo III), casos particulares del sistema generador de Dagum, la distribución de Singh-Maddala, modelo triparamétrico deducible del sistema generador de Dagum, la distribución Log-Normal,



pertenciente al sistema generador de D'Addario y las distribuciones Gamma y Beta, pertenecientes al sistema generador de Pearson.

Casi la totalidad de los trabajos realizados en las últimas décadas del siglo XX utilizaron como información muestral datos agregados, normalmente en decilas. Así, y a modo de ejemplo, Dagum utiliza los datos agregados, en 10 intervalos, relativos a la renta familiar de los EE.UU. en 1969 para demostrar la idoneidad de los modelos de Dagum – tipo I y Dagum – tipo II.

De este hecho, esto es, de la posible repercusión que la naturaleza de los datos muestrales pudiera tener en las conclusiones de las investigaciones, surge el objetivo principal del presente análisis: comprobar si, a nivel de ajuste de los modelos, los resultados obtenidos son iguales trabajando con datos agregados y con la muestra completa, es decir, con datos desagregados.

Para ello se selecciona como modelo explicativo de la renta el gamma biparamétrico. El porqué de su elección se deriva tanto de la simplicidad del modelo como por la controversia que suscitó, al ser aceptado por Salem y Mount, rechazado por Dagum y aceptado tanto por Rojo Olivas como por Lafuente.

En cuanto a la información empírica, se utilizarán los datos correspondientes a los Ingresos y Gastos familiares proporcionados por las E.P.F. de 1980-81 y 1990-91, datos que se utilizaron a finales del siglo pasado, por falta de más información, agregados en decilas. La disponibilidad de la muestra desagregada para ambos períodos permite la realización del presente estudio.

Así, y a modo de resumen, los objetivos que se abordan son:

- Analizar si es posible validar el modelo Gamma, haciendo uso como información empírica de la muestra completa.
- Comparar los resultados obtenidos, a nivel de ajuste, trabajando con los datos agregados (en decilas) y desagregados

Datos y Metodología: Tal y como se ha indicado anteriormente, los datos utilizados son los correspondientes a los Ingresos y Gastos de los hogares correspondientes a los periodos 1980-81 y 1990-91 y proporcionados por sendas E.P.F.. Conviene, en este punto, indicar que cada dato muestral (ingreso o gasto) representa a tantos hogares como indique su factor de elevación correspondiente, factor que intervendrá en la obtención de los parámetros del modelo propuesto.

En este sentido, conviene citar que dicho modelo, dependiente de los parámetros α y λ , tiene función de densidad:

$$f(x) = \frac{\lambda^\alpha e^{-\lambda x} x^{\alpha-1}}{\Gamma(\alpha)} \quad \forall x \in]0, \infty[, \alpha > 0, \lambda > 0$$

Y que el método de estimación de los parámetros utilizado será el de máxima-verosimilitud, por tratarse de un método que proporciona estimadores con buenas propiedades.

Sin embargo, dado que la aplicación de dicho método, lleva a la resolución de un sistema de ecuaciones en el que interviene la geométrica de los valores muestrales, se genera un problema al trabajar con datos agregados. En efecto, al no disponer de toda la muestra, es necesario generar algún mecanismo que permita obtener un valor aproximado de dicha media.

Respecto a los tests de adherencia, se hará uso del test de Kolmogorov-Smirnov, junto con el test que se deduce de la “verificación de las cotas de Gastwirth”, en el caso de que se trabaje en el caso de mínima información (datos agregados en decilas).

El uso de este segundo contraste se justifica en el trabajo desarrollado por Gastwirth y Smith según los cuales, para que un modelo resulte adecuado, en términos de ajuste,



debería verificar que su índice teórico de Gini pertenezca al intervalo definido por las denominadas cotas de Gastwirth, cotas que obedecen a las siguientes expresiones:

$$\text{Cota inferior: } C_I = \sum_{i \neq j} \frac{p_i p_j |\mu_i - \mu_j|}{2\mu}$$

$$\text{Cota superior: } C_S = \sum_{i=1}^A \frac{1 p_i^2 (\mu_i - e_{i-1})(e_i - \mu_i)}{\mu(e_i - e_{i-1})}$$

siendo: A el número de intervalos considerados en la agregación (10 en el caso de trabajar con decilas), p_i la proporción de hogares pertenecientes al i -ésimo intervalo, e_{i-1} y e_i los extremos que determinan el i -ésimo intervalo y μ_i los ingresos o gastos medios correspondientes a los hogares contenidos en el intervalo i -ésimo.

Análisis y resultados: con la finalidad de abordar los objetivos expuestos se considera, en un primera etapa, la situación de mínima información muestral, esto es, suponiendo que tan sólo se dispone de los datos agregados en decilas. El análisis se repite trabajando con la muestra completa y por último se comparan los resultados.

- **Análisis con datos agregados:** se parte, en este supuesto, de que la información se reduce al conocimiento de los extremos de las decilas y de los valores medios de las mismas (para Ingresos y Gastos), contrastándose, mediante los test de adherencia anteriormente citados, si cada una de las 52 provincias ajustan al modelo gamma biparamétrico para ambos periodos de tiempo.

En este caso, para obtener un valor aproximado de la media geométrica, se hace uso de la propuesta de Rojo, quien utiliza como representante de la decila i -ésima una combinación lineal convexa del extremo inferior (x_{li}) y la media (x_{mi}) de la misma, es decir: $x_{li} + k(x_{mi} - x_{li})$, donde el valor de K se determina de manera que se superen ambos test de adherencia (Kolmogorov-Smirnov y test derivado de la verificación de las cotas de Gastwirth).

Como aportación a la propuesta citada cabe indicar que se obtiene el intervalo formado por los valores de k para los que se acepta el ajuste por el test de Kolmogorov-Smirnov, verificándose al mismo tiempo las cotas de Gastwirth, así como valor de k, perteneciente al intervalo anterior, para el que se produce el mejor ajuste.

En este caso todas las provincias soportan el ajuste al modelo gamma, en ambos periodos, tanto para Ingresos como para Gastos.

- **Análisis con datos desagregados:** en este supuesto, es posible obtener la media geométrica, necesaria para la estimación de los parámetros del modelo, sin necesidad de aplicar ningún tipo de propuesta, conjetura o artificio. Recordar, en cualquier caso, que para la obtención de dicha media es necesario el uso del factor de elevación (F_i) de cada dato (D_i) correspondiente al ingreso o gastos de la i -ésima familia. Siendo $\tilde{X} = \prod_i D_i^{F_i/F_T}$, con $F_T = \sum_i F_i$

familia. Siendo $\tilde{X} = \prod_i D_i^{F_i/F_T}$, con $F_T = \sum_i F_i$

Además, se utiliza únicamente, como test de adherencia, el de Kolmogorov-Smirnov, obteniéndose que en el período 80-81 se verifica que 7 provincias (Badajoz, Barcelona, Córdoba, Huesca, Madrid, Valencia, Zamora) no superan el test de adherencia para los Ingresos y en el período 90-91 es Cuenca la única provincia que no soporta el test. Para los Gastos todas las provincias superan el test en ambos periodos.



Esto supone que, en términos porcentuales se ha producido un 96'2% de ajuste (al haberse superado el test de bondad de ajuste en 200 de los 208 contrastes realizados)

Así, se puede concluir que el modelo Gamma es válido para explicar las rentas de los hogares (medidas a través de sus Ingresos y Gastos).

Queda así cubierto uno de los objetivos propuestos: validar el modelo Gamma, haciendo uso como información empírica de la muestra completa.

- Análisis comparativo: el hecho de que se haya validado el modelo gamma biparamétrico trabajando tanto con datos agregados como con la muestra completa, permite realizar un estudio comparativo en cuanto al nivel de ajuste.

La primera conclusión que se extrae es que aunque para para ambos niveles de información la modelización propuesta es válida, el nivel de adherencia no coincide. En efecto, mientras que al considerar los datos agregados en decilas se ha producido un 100% de ajuste, dicho porcentaje es inferior (96'2%) al trabajar con los datos desagregados, repercutiendo al trabajo con Ingresos únicamente.

Así, se ha dado el caso de provincias que no soportaban el ajuste al trabajar con toda la muestra y que sí lo hacían considerando como fuente informacional las decilas muestrales.

Parece que, a falta de mejor información, en ocasiones no resultó posible rechazar la bondad del ajuste, pero si se trabajara con toda la muestra la conclusión cambiaría.

Esto hace que se los resultados de investigaciones en las que se ha utilizado este tipo de información (minimalista) deban tomarse con cautela y puedan no ser un referente adecuado para los actuales estudios de renta y desigualdad, en los que la disponibilidad de una mayor información y el auge computacional permite abarcar análisis algo más completos.

Palabras Clave: *modelización renta familiar, desigualdad regional, índice Gini, datos agregados, cotas de Gastwirth.*

Clasificación JEL: C53, E01, R29